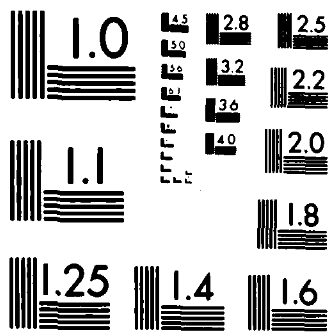
AD-A148 497 AN ANALYSIS OF ROBUST AND EFFICIENT PRIORS ASSOCIATED WITH A FINITE BAYES..(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYST.. 1/1

UNCLASSIFIED S MARINO ET AL. SEP 84 AFIT/GSM/LSY/84S-21 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

REPRODUCED AT GOVERNMENT EXPENSE

(2)

AD-A148 497

AN ANALYSIS OF ROBUST AND EFFICIENT
PRIORS ASSOCIATED WITH A FINITE
BAYESIAN MODEL FOR COMPLIANCE TESTING

THESIS

Stephen Marino, B.S. Kevin O'Shaughnessy, B.S.
Captain, USAF Captain, USAF

AFIT/GSM/LSY/84S-21

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DTIC
ELECTE
DEC 1 2 1984
S D
B

DTIC FILE COPY

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

84 12 03 140

AFIT/GSM/LSY/84S-21

AN ANALYSIS OF ROBUST AND EFFICIENT
PRIORS ASSOCIATED WITH A FINITE
BAYESIAN MODEL FOR COMPLIANCE TESTING

THESIS

Stephen Marino, B.S. Captain, USAF

Kevin O'Shaughnessy, B.S. Captain, USAF

AFIT/GSM/LSY/84S-21

Approved for public release; distribution unlimited

DTIC
ELECTE
DEC 1 2 1984
S D
B

The contents of the document are technically accurate, and no sensitive items, detrimental ideas, or deleterious information are contained therein. Furthermore, the views expressed in the document are those of the authors and do not necessarily reflect the views of the School of Systems and Logistics, the Air University, the United States Air Force, or the Department of Defense.

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

AFIT/GSM/LSY/84S-21

AN ANALYSIS OF ROBUST AND EFFICIENT PRIORS ASSOCIATED WITH A FINITE BAYESIAN MODEL FOR COMPLIANCE TESTING

THESIS

Presented to the Faculty of the School of Systems and Logistics

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Systems Management

Stephen P. Marino, B.S.

Captain, USAF

Kevin O'Shaughnessy, B.S.

Captain, USAF

September 1984

Approved for public release; distribution unlimited

Acknowledgements

Special appreciation and gratitude are given to the following individuals, without whose help this research would have never been completed:

1. Dr. James Godfrey, who developed the Finite Bayesian Procedure Model, and aided us in understanding the model's concepts while providing insight into the research design.

2. Maj Jeffrey J. Phillips, our advisor, for his guidance, criticism and motivation.

3. Ms. Jackie McHale, our expert typist, to whom we are deeply indebted.

4. Finally, to our wives, Fran and Marianne, for their support and understanding.

Table of Contents

List of Figures

List of Tables

List of Symbols

E(R**)	The expected value of the Finite Bayesian Procedure (FBP) computed minimum number of population errors. This value is computed by multiplying the hypergeometric probability of observing r errors in a sample, by the R** value associated with the particular r, and summing these values for r = 0 to r = 100.
N	The population size, set either at 1000 or 5000.
n	The sample size fixed at 100 for the study.
n'	A FBP parameter derived from the expected value and variance of the beta distribution. See Eq [16].
PROB	The hypergeometric probability of observing a certain number of errors given a population size, a sample size, and the total number of errors in the population.
R	The actual number of errors in the population.
R**	The FBP computed minimum number of population errors, based on the number of observed errors in a sample, and is equal to or greater than R, the actual number of population errors.
r**	The specific number of observed errors in the sample where R** is equal to or greater than R.
r'	A FBP parameter derived from the expected value and variance of the beta distribution. See Eq [15].
REL	The reliability of confidence level. These values represent the probability that the model results are correct. The higher the reliability, the higher the confidence for correct results.
u	The maximum tolerance error rate or materiality. For this study, u was equal to the actual number of population errors (R) divided by the population size (N).

Abstract

This research examines the reliability and validity of the Finite Bayesian Procedure (FBP) model through an evaluation of robust and efficient prior probability distributions. The model, developed by James Godfrey and Richard Andrews, presents a different approach to compliance testing in auditing.

This study utilizes small and moderate-sized populations, four population error rates, a fixed sample size, and four reliability levels. In addition, four expected error rates, based on a beta prior probability distribution and ranging from very low to high, combined with three variance levels and a uniform distribution, are used to evaluate the model.

The results indicate that the model is adequately reliable and valid. However, the uniform distribution seems to perform best of all prior probability distributions tested. Moreover, tradeoffs between robustness, efficiency, and reliability seem a necessity when using the Finite Bayesian Procedure model.

AN ANALYSIS OF ROBUST AND EFFICIENT PRIORS ASSOCIATED WITH A FINITE BAYESIAN MODEL FOR COMPLIANCE TESTING

I. Statistical Sampling in Auditing and the Finite Bayesian Procedure

Introduction

Despite somewhat staid appearances, the audit process has been dynamic in recent times. Historically, auditors often performed a 100 percent examination of the records of the organization audited before rendering an opinion as to the fairness of the financial statements (3:2). Yet, as organizations grew, and with them, the number of transactions generated, a complete review became what Arkin termed "both unwarranted and uneconomical" (3:2). Statistical sampling became a necessity for providing a reliable indication of the accuracy of other similar transactions.

Concurrent with the rise of testing or sampling, auditors began to recognize the importance of an organization's system of internal control. Essentially, auditors started depending on the internal control system, as verified by tests of that system, called compliance tests, to direct them in subsequent sampling of account balances, or substantive tests. As a general rule, the stronger the system of internal control, the less substantive testing required by the auditor (29:6).

The emergence of sampling in auditing and the rising dependence on the system of internal control provide impetus for this research effort. In 1982, Godfrey and Andrews proposed a model for compliance testing of an internal control process to determine a probability distribution for its error rate (21:304-315). The probability distribution for the error rate could then aid the auditor in his evaluation of the internal process. Their approach is Bayesian, meaning that an auditor's prior knowledge and experience are statistically incorporated into the sampling technique. Also, the model assumes finite populations, which correctly describe those encountered in auditing situations. When compared to other common methods of compliance testing, Godfrey and Andrews propose that their model, called the Finite Bayesian Procedure (FBP), more closely emulates the actual auditing environment and requires sample sizes equal to or less than those required by the common alternatives, such as the classical method or the Bayesian procedure introduced by Felix and Grimlund (21:304-305).

Research Objectives

This study expands on Godfrey and Andrews' analytical work by testing the model's effectiveness in evaluating the degree of compliance with an internal control process. The research centers on the evaluation phase of the FBP with two primary objectives:

1. To determine the model's reliability and validity in a typical auditing environment;

2. To determine if any specific prior distributions significantly improve the model's reliability and validity.

Background

An Overview of the Auditing Process.

> Auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users [11:18].

Depending on the perspective, the purpose of the audit varies. For the independent public accountant, whose view this report takes, the ultimate objective of the examination is the expression of an opinion on the fairness with which the financial statements present the financial position of the organization audited in accordance with generally accepted accounting principles (2:5). For the internal auditor, the concern is an examination and appraisal of the internal control system's integrity and the efficiency of financial, accounting, and administrative operations, to assure compliance with established procedures and to provide the basis for improvement in operations (2:5). The governmental auditor's goal is similar to that of both the independent public accountant and the internal auditor, but also includes "auditing the economy, efficiency, and achievement of desired results" (12:2). Though the final objectives vary somewhat, each auditor must evaluate the actual degree of compliance with the organization's system of internal control.

The study and evaluation of the system of internal control provides the starting point for independent auditors in defining the nature, timing, and extent of subsequent audit procedures required for the expression of an opinion on financial statements (25:10-3). Statements on Accounting Standards formally defines "internal control":

> Internal control comprises the plan of organization and all of the coordinate methods and measures adopted within a business to safeguard its assets, check the accuracy and reliability of its accounting data, promote operational efficiency, and encourage adherence to prescribed managerial policies. . . . [A] "system" of internal control extends beyond those matters which relate directly to the functions of the accounting and financial departments [1:Section 320.09].

This broad definition encompasses both internal accounting control and administrative, or operational, control, though the distinction is often not clear cut.

Administrative or operational controls refer to the plans, methods, and measures used to provide operational efficiency and adherence to prescribed policies in all departments of the organization (25:10-7). Generally, controls of this nature do not bear directly on the financial statements and, consequently, probably do not directly interest the independent auditor.

Internal accounting controls, on the other hand, do directly interest the independent auditor because they bear directly upon the dependability of the accounting records and the financial statements. SAS No. 1 defines "internal accounting controls" as:

> the plan of organization and the procedures and records that are concerned with the safeguarding of assets and the reliability of financial records and consequently are designed to provide reasonable assurance that:
>
> a. Transactions are executed in accordance with managements' general or specific authorization.
>
> b. Transactions are recorded as necessary (1) to permit preparation of financial statements in conformity with generally accepted accounting principles or any other criteria applicable to such statements and (2) to maintain accountability for assets.
>
> c. Access to assets is permitted only in accordance with management's authority.
>
> d. The recorded accountability for assets is compared with the existing assets at reasonable intervals and appropriate action is taken with respect to any differences [1:Section 320.28].

The independent public accountant's examination focuses on these accounting controls.

The examination extends beyond defining and reviewing an internal control system. Before deciding on the nature and extent of subsequent auditing procedures, the auditor must gain some assurance of the system's protective quality (3:3). Superficially, the internal control system may appear excellent, but given the requirement for objective evidence on which to base the auditor's judgment, an evaluation of the degree to which the system operates as prescribed is needed. Errors may occur deliberately, such as by fraud or embezzlement, or inadvertently, by random clerical failures (3:3). Also, they may arise in areas not adequately protected by the internal control system (3:3). Only by examining records processed through the system can the auditor gain assurance that it operates effectively. A sample, rather than a

complete inspection of all records, may serve as the basis for the evaluation. These samples are called "tests of transactions," or more commonly, "tests of compliance."

Arkin notes that,

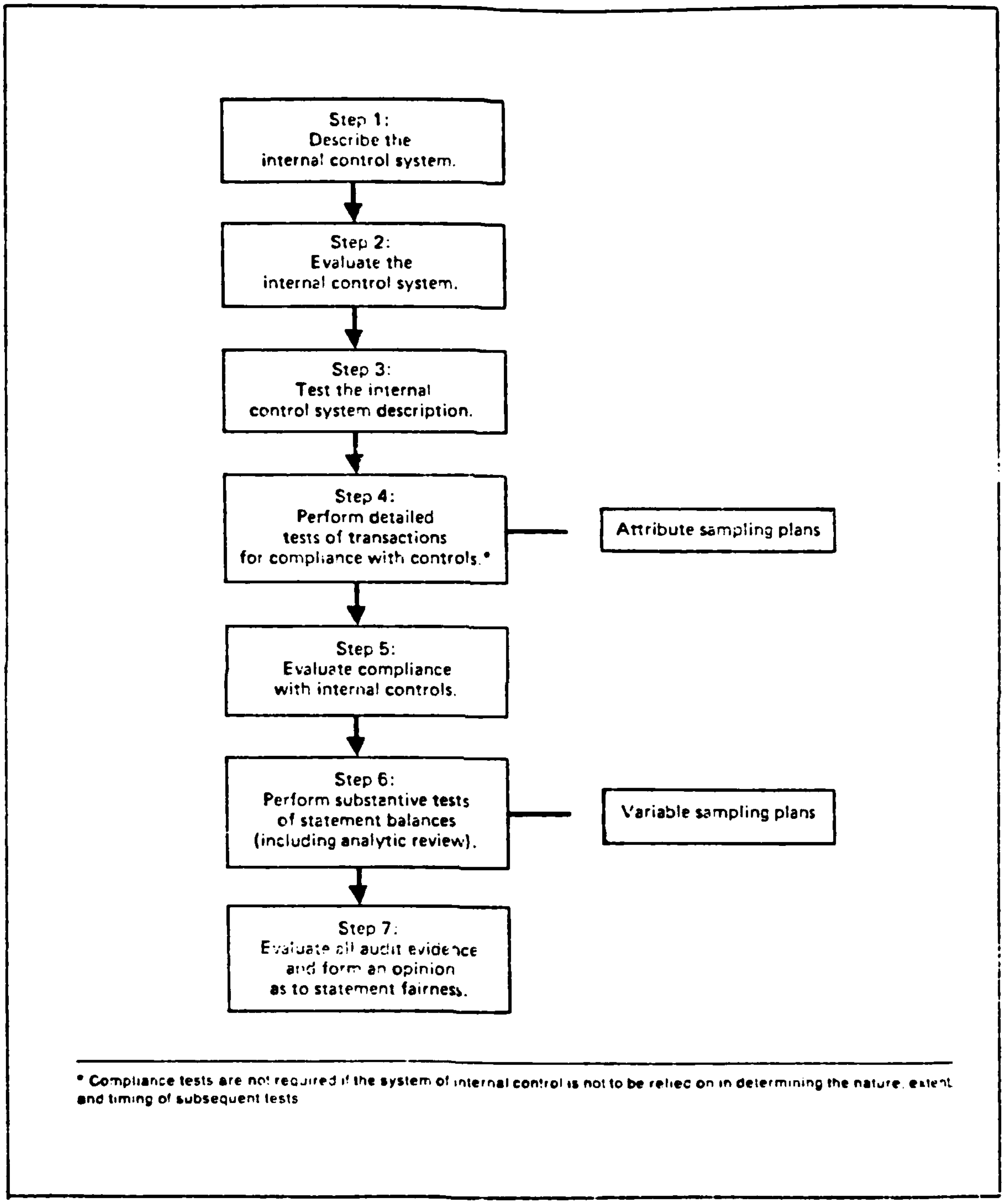
> Apart from the possible failure or inadequacy of the internal control system is the auditor's direct determination of the accuracy and reliability of the values that appear in the financial statements [3:4].

Again, auditors typically perform tests. These tests of account balances are referred to as "substantive tests."

The audit process, therefore, is integrated and multi-phased and can involve numerous tests, both compliance and substantive. Kenneth P. Johnson summarizes the approach:

> Under the profession's *standards, auditors first* review a company's system of internal accounting control and make an evaluation of it. Based on the evaluation, they determine whether to perform functional (compliance) tests of internal accounting controls, or to proceed directly to the validation (substantive) testing of account balances. If the evaluation of internal accounting controls indicates that a control procedure has been established, the auditor has the option of functionally testing the control to gain reasonable assurance that the procedure is in effect, is operating as prescribed, and can be expected to continue to do so throughout the period under examination, and thus can appropriately reduce the validation testing of related account balances. If, on the other hand, the evaluation uncovers a control weakness, the auditor cannot rely on the control and has to select validation tests of the appropriate nature, extent, and timing to compensate for the control weakness, and apply them to the related account balances [25:10-3].

Figure 1 provides a simplified view of the audit process and illustrates the stages when sampling techniques may prove useful to the auditor.



(Adopted from 4:6)

Figure 1. The Audit Process

The Role of Sampling in Auditing. Due to the tremendous amount of transactions processed through modern organizations, auditors must rely on partial examinations or samples on which to base their evaluation. Meigs, et al., defines sampling as "the process of selecting a sample from a larger field of items (called the population) and using the characteristics of the sample to draw inferences about the characteristics of the entire population" (29:238). Sampling can take two basic forms: judgmental and statistical. In judgmental sampling, the auditor uses his discretion in predetermining the size and composition of each audit sample (29:227). Bailey notes, "Convenience, availability, whim, or any other factor may be applied" (4:30). Judgmental sampling provides no means for verifying that the sample represents the population, or for measuring the possibility that it does not (29:239). On the other hand, with statistical sampling the auditor selects desired levels of risk and precision and estimates the expected number of exceptions for each sample. He then refers to statistical sampling tables to determine the appropriate sample size and, in some cases, the specific transactions for inclusion in the sample (29:227).

Because statistical sampling involves randomness, where every unit of the population has an equal chance of selection, it offers several distinct advantages. Most notably, statistical sampling enables measurement of the risk of material errors occurring purely by chance and allows justifiably smaller sample sizes -- both of which are not possible with judgmental sampling (29:329, 268-269).

Statistical sampling refers to a broad classification of the many statistical methods available to the auditor. Two basic types of methods separate this classification: 1) attribute sampling plans, normally associated with compliance tests; and 2) variable sampling plans, usually relating to substantive tests. When conducting attribute tests, concern lies with the qualitative aspects of sample units, usually expressed in dichotomous groupings (4:47). These are tests of the YES/NO form, such as error or no error, authorized or not authorized, and so on (4:47). Variable sampling plans address the quantitative aspects or values of the sample units, usually in dollar amounts (4:47). Normally, the concern rests with estimating the dollar value of errors associated with the sampled account balances. This study focuses on a procedure for compliance testing using an attribute sampling plan.

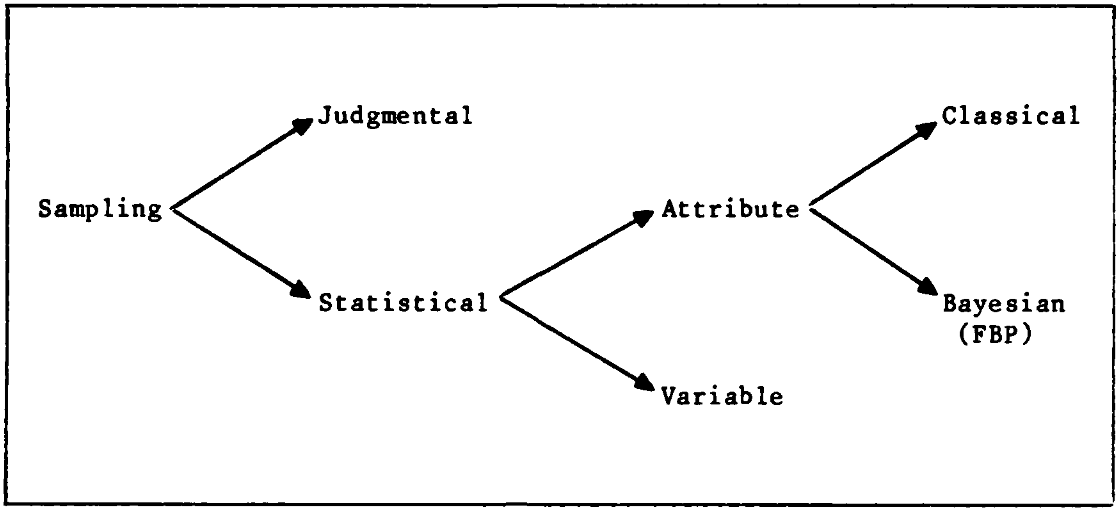
Auditors have several methods of attribute sampling available to them. Traditionally, these techniques are of the classical statistics variety; however, research over the past several years suggests a separate classification for attribute sampling using Bayesian procedures. The distinction seems especially appropriate for the purposes of this research effort, which evaluates a Bayesian procedure proposed by Godfrey and Andrews.

Classical procedures use techniques derived from classical statistics similar to those methods used widely in business, industry, and science. Briefly, the general procedure is as follows:

After specifying a desired precision (reliability) and confidence level (risk) for the sample estimate, the auditor makes a preliminary estimate of the fraction or percent in error in the population. The auditor may derive this estimate from previous data or through a preliminary sample, or pilot study. By referencing various precomputed tables readily available, the auditor then estimates the initial sample size. He then randomly draws this sample and computes the fraction in error. Iteratively, he repeats the process until attaining the desired precision and confidence (i.e., the auditor re-determines the required sample size, randomly draws this redetermined sample, and recomputes the fraction in error until the error rate falls within the specified reliability and risk parameters). Alternatively, the auditor may reevaluate and relax the specified precision and confidence levels to preclude additional sampling [16:208-211; 29:253-260].

Apparently, several weaknesses exist in the use of classical procedures for auditing. As a practical matter, the auditor may find it difficult, if not impossible, to conduct the additional sampling often required by the classical procedure. The auditor normally must relax the specified reliability and/or confidence level when the sampling results so warrant, forcing an ad hoc evaluation (29:269-270). Furthermore, auditors often find it difficult to incorporate the results of the classical procedure, usually expressed as confidence intervals or tests of significance, into final subjective judgments concerning the quality of the organization's internal control system (32:265).

Bayesian procedures have emerged in response to problems with the classical approach. The Bayesian method incorporates a subjective assessment of the internal control process into the analysis. The procedure recognizes that the auditor may have a significant amount of experience which can aid in making estimates and in selecting sample sizes, that the auditor considers certain qualitative factors in making



(Adopted from 27:2)

Figure 2. Relationship of FBP to Sampling Classifications

his estimates, and that sampling is costly in both time and money (5:115). Using Bayes Theorem, this approach provides a methodology for explicitly including subjective assessments into the analysis, and thereby provides a framework for the ultimate decision on the quality of the organization's system of internal control. Figure 2 illustrates the relationships and positions this study of a Bayesian approach among the various sampling methods discussed thus far.

The Bayesian Approach: An Overview. The potential advantages of a Bayesian approach relating sample evidence to auxiliary information have been recognized in the accounting literature for some time. The pioneering efforts by Birnberg and by Kraft highlighted not only the possible benefits in sample size and sample design, and the associated reduced cost, but also the significance of formally incorporating the auditor's intuition into the decision process (5:108-116; 26:50). By

recognizing and quantifying this prior information and by providing a methodology for explicitly including it in the decision process, the Bayesian approach tends to prevent haphazard use of this experience (5:114).

Several researchers, including Tracy, Sorensen, and Godfrey and Andrews, emphasize the importance of this explicit incorporation of subjective evidence, comparing it to classical statistical inference procedures (34; 33; 21). The Bayesian view explicitly weighs, in the form of a prior distribution, the auditor's beliefs concerning the probability of events (e.g., population error rates). On the other hand, the classical procedure implicitly assumes equal probability for various population error rates. A uniform distribution of errors seems highly unrealistic, however, in the typical low error rate audit environment (34:93; 33:555-558; 21:305). Scott, Ijiri and Leitch, and others cite legal reasons favoring the Bayesian approach as well (31; 24:106).

Sorensen explains the Bayesian framework for discrete variables:

> In general terms, the Bayesian theorem is a given set of mutually exclusive and collectively exhaustive events $E_1, E_2, \ldots, E_n$, and an experimental outcome, e, such that
>
> $$P(E_j|e) = \frac{P(e|E_j)\,P(E_j)}{\sum_{i=1}^{n} P(e|E_j)\,P(E_j)} \qquad [1]$$
>
> for $j = 1, 2, 3, \ldots n$

where

$P(E_j|e)$ = posterior probability of E_j being the state of the world, given experimental outcome, e

$P(e|E_j)$ = conditional probability of experimental outcome, e, given E_j is the state of the world

$P(E_j)$ = prior probability on E_j being the state of the world

$P(e|E_j)\ P(E_j)$ = joint probability

$\sum P(e|E_j)\ P(E_j)$ = marginal probability of experimental outcome, e, summed over all possible states of the world [33:557]†

† A more rigorous description of the general Bayesian approach for the continuous case can be found in Godfrey and Neter's report in the Audit Research Working Paper Series (Report No. 83-001), entitled "Bayesian Bounds for Monetary-Unit Sampling in Accounting and Auditing," University of Georgia, 1983.

Two assumptions concerning the prior probabilities are often made throughout the literature relating to attribute sampling. First, it is assumed that auditors can use their judgment and other evidence to formulate specific probabilistic estimates of the population error rate (e.g., $P(E_j)$ (19:26). Crosby, Chesley, Felix, Corless, and others have demonstrated in assessment studies that this elicitation is possible, though much variability exists (7-10; 13; 14; 15; 18). Furthermore, the particular elicitation technique used appears to play a key role in the obtained probability, though training seems to improve the confidence in these assessments (14:356). To date, the research evidence is inconclusive in assessment of prior probabilities.

The second assumption concerns the case of a beta-binomial model. Blocker, Corless, Crosby, Felix and Grimlund, and others have used the beta family of probabilities to represent the auditor's prior beliefs (6; 13; 14; 15; 19). Though initially chosen for its reasonableness and analytic simplicity, Crosby demonstrated that the beta probability distribution did adequately model the auditor's beliefs, indicating its descriptive validity as well (14:364).

The beta-binomial model also assumes that the underlying process from which the sample is taken (e.g., some internal control procedure) can be represented as a Bernoulli process. As a result, the binomial probability distribution expresses the likelihood of finding r errors in a sample of size n (15:588). Crosby notes:

> A Bernoulli process involves independent samples with replacement or sampling from an infinite population, so that the probability of selecting an error transaction remains constant from trial to trial. However, such is not the case in auditing. In general, sampling in an audit context is without replacement from a finite population. The formula analogous to the binomial probability function which applies to sampling without replacement, in which successive trials are not independent, is the hypergeometric distribution. [15:588-589].

<u>The Finite Bayesian Procedure</u>. Godfrey and Andrews recently proposed a Finite Bayesian Procedure (FBP) for compliance testing using the hypergeometric distribution. Inspired by Ericson's work on applying Bayesian procedures to finite population sampling, the Godfrey and Andrews model improves on the Infinite Bayesian Procedure (IBP) introduced by Felix and Grimlund (17; 19). The FBP correctly assumes a finite population in the auditing environment and thus, more closely emulates that environment than either the classical method or the IBP

(21:304-305). Furthermore, for typical auditing populations, the FBP requires smaller sample sizes than the classical procedure and sample sizes less than or equal to those required by the IBP (21:305).

Godfrey and Andrews take a three-phased approach toward making a conclusion on the finite population error rate, or the percentage of time the internal control process erroneously processes a given attribute in the population. This three-phased approach involves planning, random sampling, and evaluation (21). In the planning phase, the procedure assesses the auditor's prior information and judgment concerning the probable distribution of errors and subsequently uses this information for determining the required sample size, relative to specified parameters. In random sampling, the procedure directs the use of some random sampling plan to draw the required sample size. In the evaluation phase, the model incorporates the prior distribution and sample evidence to make a statistical statement on the population process error rate. This statement then provides the auditor with objective evidence on which to base his ultimate judgment on the quality of the internal control process.

Though mathematically complex, a brief outline of the three-phased procedure follows:

<u>Planning</u>.

1. Specify the desired parameters for the sampling plan.

    A. Population size (N)

    B. Reliability (REL*) or confidence level

    C. Maximum Tolerance Error Rate (u)

Based on auditor judgment, set u such that if the true population

error rate exceeds u, the attribute tested would be considered out of control. The specified u represents the materiality, or upper precision limit. This limit is the maximum error rate the auditor will permit and still rely on the controls of the system. The value u is transformed to reflect R*, the maximum tolerable number of errors in a finite population of size N, such that

$$R^*/N \leq u \qquad [2]$$

2. Assess the prior distribution on p, the process error rate. Based on the auditor's prior information, an elicitation method derives a prior distribution on the anticipated error rate. The technique requires the auditor to directly and subjectively judge the probability of zero errors, (P (r = 0)), and the probability of one error, (P (r = 1)), based on an arbitrary sample size, n (e.g., n = 100). From the direct elicitation or assessment of these probabilities, the prior distribution, assumed to be a beta distribution, is indirectly assessed by an iterative procedure, solving for parameters r' and n', using:

$$P(r) = \binom{n}{r} \frac{\Gamma(r' + r)\,\Gamma(n')\,\Gamma(n' - r' + n - r)}{\Gamma(r')\,\Gamma(n' + n)\,\Gamma(n' - r')} \qquad [3]$$

where

$\Gamma(\alpha)$ is the gamma function given by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha - 1} e^{-t} dt, \quad \alpha > 0 \qquad [4]$$

Essentially, the two values, P (r = 0) and P (r = 1), define the prior distribution mathematically transformed into values for r' and n'.

The prior distribution on p, the process error rate, is then fully described in terms of parameters r' and n', such that

$$E(p) = r'|n' \qquad [5]$$

$$\sigma^2(p) = \frac{r'(n' - r')}{n'^2(n' + 1)} \qquad [6]$$

3. Determine the minimum required sample size, n. Based on the previously specified values and assessed prior parameters, and using

$$P''(R|r) = \binom{N - n}{R - r} \frac{\Gamma(r' + R)\,\Gamma(n' + n)\,\Gamma(n' - r' + N - R)}{\Gamma(r' + r)\,\Gamma(n' + N)\,\Gamma(n' - r' + n - r)} \qquad [7]$$

for r = r*,

where

r* is the maximum number of errors allowed in the sample for acceptance of the internal control procedures,

solve for the minimum n such that

$$\sum_{R = r^*}^{R^*} P''(R|r = r^*) \geq REL^* \qquad [8]$$

Random Sampling.

4. Randomly draw the required sample size, n, and count the number of errors, r, in the sample.

Evaluating.

5. Given the sample data, determine if the control procedure is acceptable using the decision rule:

$$\sum_{R = r}^{R^*} P''(R|r) \geq REL^*$$

where

$$P''(R|r) = \binom{N - n}{R - r} \cdot \frac{\Gamma(r' + R)\,\Gamma(n' + n)\,\Gamma(n' - r' + N - R)}{\Gamma(r' + r)\,\Gamma(n' + N)\,\Gamma(n' - r' + n - r)} \qquad [9]$$

for $R = r, r + 1, \ldots, N - n + r$

Essentially, this rule states that, given the planned sample size and specified confidence level, if the observed number of errors, r, is less than the maximum allowable number of errors in the sample, r*, (i.e., $r < r^*$), then the internal control procedure will be accepted (21:308-309).

The data from the planning and sampling phases determine all the values required for these later computations. The conclusion of this evaluation phase provides a statistical statement on the process error rate which can aid the auditor in the ultimate decision on the quality of the internal control process.

Research Questions

The complexity of the FBP requires a computer software package or tabulated values to afford any practical significance to the auditor. These conveniences have not yet been developed. However, before doing

so, the validity of this procedure warrants testing. This research analyzes the effectiveness of the FBP, centering on the evaluation phase of the procedure. Specifically, the research questions address:

1. How reliable and valid is the FBP at estimating population error rates in typical audit environments?

2. Are there any robust and efficient prior distributions that significantly improve the model's reliability and validity?

II. Research Design

Overview

The essence of the FBP is the decision rule it provides auditors for determining whether to accept or reject the particular population under examination. This decision rule results from the inferential procedure which occurs in the evaluation phase, given by Eq [9], wherein a statistical statement on the population process error rate is made to guide the auditor in the ultimate judgment on the quality of the internal control process. This study assumes the previous accomplishment of the planning and sampling phases of the procedure and concentrates, rather, on the evaluation phase, specifically targeting the reliability and validity of the inferential procedure.

Intrinsic to the inferential procedure are the parameters r' and n' which define the specific prior probability distribution associated with the particular population under investigation. As previously discussed, these "priors" represent the auditor's beliefs, based on past experience and judgment, concerning the expected error rate and its variance. The particular prior probability distribution chosen by the auditor influences the posterior probability distribution and, hence, the decision on accepting or rejecting the internal control procedure. It seems reasonable, therefore, to search over several different priors to find those that are "robust" and "efficient" for a given set of audit populations.

Robustness. The concept of "robustness" is linked with reliability and validity. As Huber notes, "robustness signifies insensitivity to small deviations from the assumptions" (23:1). This insensitivity directly influences the ability of a model to approximate the true underlying situation. Huber elaborates on the significance of robustness, stating:

> Therefore, any statistical procedure should possess the following desirable features:
>
> (1) It should have a reasonably good (optimal or nearly optimal) efficiency at the assumed model.
>
> (2) It should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly, that is, the latter . . . should be close to the nominal value calculated at the model.
>
> (3) Somewhat larger deviations from the model should not cause a catastrophe [23:5].

As used in this study, robustness involves the insensitivity of the model's inferential procedure to slight variations in the population from the assumed prior probability distribution. A robust prior distribution provides coverage, such that its influence on the posterior probability distribution is correct for any actual population encountered. That is, though the actual population error rate may differ substantially from the expected one, a robust prior probability distribution correctly provides, through the model, a bound on the population error rate that equals or exceeds the actual population error rate. This bound represents the minimum error rate in the population for the specified reliability and given number of observed errors in the sample.

From the auditor's perspective, denoting those priors which provide coverage as robust seems appropriate. The auditor's concern lies with estimating the probability that the population under investigation has an error rate that exceeds some predetermined materiality, u, with reasonable assurance of the results. In practice, the actual number of population errors may not be known. The FBP aims at estimating the true, but unknown, population error rate through a probabilistic statement, such that the likelihood of an actual population rate exceeding the computed error rate is less than or equal to 1 - REL (i.e., α). Mathematically, then, acceptable priors are those for which

$$P(R > R^{**}) \leq 1 - REL \qquad [10]$$

where

R = the total number of population errors

R** = the FBP computed bound

REL = specified reliability

The determination of robust priors provides a general indication of the model's reliability and validity. For example, if the selected prior distribution describes an expected error rate of 10 percent and 10 errors are observed in a sample of 100 (resulting in a bound approximating a 10 percent estimated population error rate), the actual population error rate should be no greater than 10 percent for the model to be considered reliable and valid. If the actual population error rate exceeds 10 percent with a probability exceeding 1-REL, then the model is not providing coverage and its reliability and validity

are suspect. For the purposes of this study, the number of robust priors served as a general barometer of model reliability and validity.

From Godfrey and Neter's discussion of robustness one can infer its significance to the usefulness of the model: "A robust . . . [prior probability distribution] . . . could be used by auditors with confidence even when they have doubts about the nature of the prior distribution" (22:27). Thus, though an auditor may feel quite uncertain about the correctness of the assumed population error rate, expressed as a prior distribution, he could confidently use a robust prior distribution with reasonable assurance that it would provide a correct bound on the actual population error rate.

Efficiency. Equally important as robustness is "efficiency." Efficiency involves the tightness or closeness of the bound to the actual population error rate. Tradeoffs between efficiency and robustness may be necessary. For example, the most robust prior may provide coverage for all anticipated populations, but may drastically overestimate the true population error rate. A prior distribution with a 90 percent expected error rate, for example, would cover most all populations possibly encountered in the audit environment. Yet, even with a relatively low error sample outcome, the computed error rate, or bound, would probably severely overestimate the true error rate. Efficiency, therefore, should be balanced with robustness.

Three-Phased Methodology. This study focuses on an analysis of various prior distributions to determine those that are robust and efficient for a given set of audit populations. The approach follows three phases:

1. Determination of those prior distributions compatible with the beta distribution assumed in the model, and the selection of test values;
2. Determination of those compatible prior distributions that are robust;
3. Determination of those robust prior distributions that are efficient.

Phase 1: Determining Compatible Priors and Selection of Test Values

Analyzing Compatibility with the Beta Distribution. The FBP assumes the beta distribution models both the prior and the posterior probability distributions. As noted earlier, Crosby showed the descriptive validity of the beta distribution in modeling auditors' beliefs (14:364). Further, the beta distribution represents a "conjugate prior" with respect to the Bernoulli likelihood function, which models the dichotomous situation faced in the compliance testing process (35:148-149). The beta distribution, however, imposes some limitations on the selection of priors by auditors.

Winkler explains the beta distribution as follows:

> Generally, the beta distribution on p (the process error rate, in this case) with parameters r and n, where $n > r > 0$, is given by

$$f(p) = \frac{(n-1)!}{(r-1)!\,(n-r-1)!} p^{r-1} (1-p)^{n-r-1} \qquad [11]$$

for $0 \leq p \leq 1$

elsewhere, 0

This represents a continuous distribution where parameters r and n need not be integers. However, if r and n are not integers, the factorial terms (n - 1)!, (r - 1)!, and (n - r - 1)! must be replaced by gamma functions, denoted by $\Gamma(n)$, $\Gamma(r)$, and $\Gamma(n - r)$. Formally,

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx, \text{ for } t > 0 \qquad [12]$$

For integer values of t,

$$\Gamma(t) = (t - 1)! \qquad [13]$$

The mean and the variance of a beta distribution with parameters r and n are:

$$E(p|r,n) = \frac{r}{n} \qquad [14]$$

$$\sigma^2(p|r,n) = \frac{r(n - r)}{n^2(n + 1)} \quad [35:149\text{-}150] \qquad [15]$$

The parameters r' and n', as used in the FBP, can be expressed in terms of the mean and the variance as follows:

$$r' = \left[\frac{(1 - E(p))\,E(p)}{\sigma^2(p)} - 1\right] E(p) \qquad [16]$$

$$n' = \frac{(1 - E(p))\,E(p)}{\sigma^2(p)} - 1 \qquad [17]$$

The limitations on the parameters, where $n' > r' > 0$, imply that

1.) $E(p) > \sigma^2(p)$ [18]

2.) $E(p)\,(1 - E(p)) > \sigma^2(p)$ [19]

Generally, these requirements for selecting "compatible" priors do not impose serious constraints on the model, but do demand some vigilance in the selection of prior probability distributions. The various prior distributions chosen for analysis in this study were tested for compatibility before evaluation.

Selection of Priors for Evaluation. Efforts were made to select a wide range of priors which adequately represent typical audit situations. Under the assumption that the normal audit environment involves relatively few errors, four corresponding expected error rates were chosen. These values represent a range of error rates from very low (.005), to low (.01), to moderate (.05), to high (.10). Error rates in excess of .10 would indicate a defective internal control process, on which the auditor could not rely. In such cases, auditors would likely proceed directly to substantive testing (20).

Defining a particular prior distribution requires the specification of variance, a measure of dispersion, in addition to the expected error rate. Generally, if a great deal of homogeneity in the data exists, and thus, small deviations from the expected error rate, then the corresponding variance is small. The specific variances studied were determined using the coefficient of variation, C, such that

$$C = \frac{\sigma(p)}{E(p)} \qquad [20]$$

where

$\sigma(p)$ is the standard deviation, or the square root of the variance.

The coefficient of variation is a dimensionless statistic useful for comparing the dispersions of two or more different types of data (28:211-212). Three values for the coefficient of variation were selected for each error rate analyzed, resulting in a large variance ($C = 2$), a moderate sized variance ($C = 1$), and a small variance ($C = .10$).

Table I summarizes the thirteen prior distributions investigated in this study. The first twelve depict those previously discussed, and the last, where $r' = 1$ and $n' = 2$, describes a uniform distribution. This uniform distribution defines the non-informative prior judgement implicitly assumed in classical techniques and provides a means for comparison of those techniques with the FBP.

Selection of Sample Size, Test Populations, and Reliability Levels. Each of the thirteen selected prior distributions were individually tested against a variety of assumed populations at several reliability levels using a constant sample size. As with the selection of priors, the values chosen for test populations and reliability emphasized what a practitioner could possibly encounter in audit situations. The selected values correspond to variables in Eq [9] and fall within four categories: 1) sample size; 2) population size; 3) actual population error rate; and 4) reliability.

Sample Size (n). Fixing sample size at $n = 100$ throughout the study limited the computational requirements of the research. This single value follows the assumption mentioned earlier concerning the previous completion of the planning phase. In practice, the planning phase of the FBP would dictate the actual sample size required to

TABLE I

Prior Parameters Selected for Analysis

E(p)	Coefficient of Variation	$\sigma(p)$	$\sigma^2(p)$	Priors r'	Priors n'
.005	2.0	.0100	.00010000	.24375	48.75
.005	1.0	.0050	.00002500	.99000	198.00
.005	0.1	.0005	.00000025	99.49500	19899.00
.010	2.0	.0200	.00040000	.23750	23.75
.010	1.0	.0100	.00010000	.98000	98.00
.010	0.1	.0010	.00000100	98.99000	9899.00
.050	2.0	.1000	.01000000	.18750	3.75
.050	1.0	.0500	.00250000	.90000	18.00
.050	0.1	.0050	.00002500	94.95000	1899.00
.100	2.0	.2000	.04000000	.12500	1.25
.100	1.0	.1000	.01000000	.80000	8.00
.100	0.1	.0100	.00010000	89.90000	899.00
.500	0.577735	.288675	.08333333	1.00000	2.00

accomplish the statistical objective, based on the specified prior distribution. Here, the selection of $n = 100$ provides a sample large enough to adequately represent the population and coincides with the samples used in comparable studies, such as by Felix and Grimlund (19:34-35).

Population Size (N). Two population sizes were selected: 1) $N = 1000$, representative of a relatively small population, and 2) $N = 5000$, representative of a moderate-sized population.

Actual Population Error Rate (u). Four population error rates were chosen, representing limits on the materiality, u, and include very low (.005), low (.01), moderate (.05), and high (.10) error rates. The values match those chosen for expected error rates expressed in the prior distributions. Using $u = R/N$, as in Eq [2], the selected error rates can be shown as actual population errors, R, for the two population sizes.

Reliability (REL). The selected reliabilities, or confidence levels, ranged from lower confidence (.85) to moderate confidence (.90 and .95) to high confidence (.99). These values represent the various probabilities that the model results are correct (i.e., that the actual number of errors in the population are equal to or less than the amount the model predicts). The higher the reliability, the more confidence the auditor can have in the results. Formally,

$$REL = 1 - \alpha \qquad [21]$$

where α represents the probability of committing a Type I error. In this study, a Type I error implies accepting the population as having

an error rate equal to or less than the materiality level, u, when it actually contains more errors than the specified value.

Table II summarizes the various populations and reliabilities against which each prior distribution was tested. The diverse range of test values encompasses those the auditor might encounter when using the FBP. The research results, then, should indicate the reliability and validity of the model.

Phase 2: Determining Robust Priors

For each of the test populations specified in Phase 1, the FBP results were compared to those of the hypergeometric distribution to determine robust prior distributions. These robust priors show the reliability and validity of the model and provide a stepping stone for subsequent determination of efficient prior distributions.

The Hypergeometric Distribution as Comparative Instrument.

Determining the robustness of prior distributions required comparison with accounting populations. Three methods were initially considered. First, actual populations with identified errors could have been employed, but data was not readily available. Second, simulated populations with randomly distributed errors could have been used. Other studies used this technique, such as those by Felix and Grimlund and by Neter and Loebbecke, among others (19; 30). The third method involves a direct determination of the probability of sample outcomes, based on a specified error rate, using the hypergeometric distribution. This technique provides more accurate results than a simulation study and enables broader generalization than a study of individual populations.

TABLE II

Summary of Test Populations and Reliability Levels

N = 1000, n = 100			N = 5000, n = 100		
u	R	REL	u	R	REL
.005	5	.85	.005	25	.85
		.90			.90
		.95			.95
		.99			.99
.01	10	.85	.01	50	.85
		.90			.90
		.95			.95
		.99			.99
.05	50	.85	.05	250	.85
		.90			.90
		.95			.95
		.99			.99
.10	100	.85	.10	500	.85
		.90			.90
		.95			.95
		.99			.99

The hypergeometric distribution models the compliance testing process, using random sampling from a finite population without replacement. This conditional distribution represents the probability of observing r errors in a sample of size n from a population of size N containing R errors, given by:

$$P(r|n,N,R) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}} \qquad [22]$$

The hypergeometric distribution was used to measure model reliability. The exact probabilities of observing r errors in a sample of n = 100 were computed, conditional on the assumed populations given by N and R as discussed in Phase 1 (i.e., $P(r = 0|n,N,R)$, $P(r = 1|n, N, R)$, . . . , $P(r = 100)|n,N,R)$. These probabilities varied across populations. For any particular population, however, a specific sum of these hypergeometric probabilities was compared with the designated reliability to determine the robustness of the priors.

Procedure for Determining Robust Priors. The procedure followed three steps for each prior distribution.

Step 1. Per the FBP, determine the minimum number of errors in the population, R**, based on the number of errors observed in the sample, r. Initially, the observed number of errors incremented from r = 0 to r = 100 and, with each iteration, the following inequality was computed:

$P(R \leq R^{**} | r', n', N, r, n) =$

$$\sum_{R = r}^{R^{**}} \binom{N - n}{R - r} \frac{\Gamma(r' + R)\,\Gamma(n' + n)\,\Gamma(n' - r' + N - R)}{\Gamma(r' + r)\,\Gamma(n' + N)\,\Gamma(n' - r' + n - r)} \geq \text{REL} \qquad [23]$$

where

R^{**} represents the minimum value of R for which the inequality holds.

This first step provides, for each value of r, the minimum number of errors in the population, per the FBP, at the designated reliability level.

Step 2. Determine the values of r for which $R^{**} \geq R$. After computing the values of R^{**} for a given prior distribution, these values were then compared to the assumed number of errors, R, for that particular population size, N. The minimum R^{**} that equals or exceeds R provides a bound for this population, and is based on a specific number of observed errors in the sample, r^{**}. The value for r^{**} represents the minimum number of observed errors required to establish this bound (i.e., where min $R^{**} \geq R$). Generally, as r increases from $r = r^{**}$ to $r = 100$, the maximum possible number of observed errors, the bound increases (i.e., the computed number of population errors increasingly exceeds the actual number of population errors). The value of r^{**}, therefore, is directly associated with the bound, R^{**}, and those values of r from $r = r^{**}$ to $r = 100$ specify the realm in which the model's computed number of errors equals or exceeds the actual number of population errors.

Step 3. Determine the hypergeometric probability of the sample outcome and compare to the specified reliability. Given the value of r** for a particular test population, the hypergeometric probabilities associated with that population were summed over the range from r = r** to r = 100, as shown:

$$P(r \geq r^{**} | n, N, R) = \sum_{r = r^{**}}^{100} P(r) \qquad [24]$$

The resultant probability represents the likelihood of sample outcomes where the model provides coverage, or, when subtracted from one, the probability that the model does not provide coverage.

This hypergeometric probability, when compared with the specified reliability level, determines the robustness of the prior distribution. Specifically, if

$$\sum_{r = r^{**}}^{100} P(r) > REL \qquad [25]$$

or, equivalently, if

$$\sum_{r = 0}^{r^{**} - 1} P(r) \leq 1 - REL \qquad [26]$$

then the specified priors, r', n', were considered robust for the population under consideration.

Phase 3: Determining the Efficiency of Robust Priors

After analyzing the various prior distributions for robustness, those identified as robust were evaluated for efficiency. As discussed earlier, efficiency measures the tightness of the bound to the actual

number of population errors. To enable comparison between population, a normalized measure was devised based on the expected bound.

The expected bound, E(R**), was computed by multiplying the hypergeometric probability of observing r errors in a sample, P(r), by the R** value associated with the particular r, and summing these values for r = 0 to r = 100. That is,

$$E(R^{**}) = \sum_{i=0}^{100} R_i^{**} P(r = r_i) \qquad [27]$$

where

$0 \leq \text{Efficiency} \leq \infty$

This value provides the expected number of errors in the population computed by the FBP, for given values of r', n', and N.

The numer of errors by which the FBP overestimated the actual number of population errors was determined by subtracting the actual numer of errors, R, from the expected bound, E(R**). Dividing this value by the actual number of population errors resulted in the tightness of the bound, expressed as a percentage of actual population errors. The measure of efficiency, therefore, is given by:

$$\text{Efficiency} = \frac{E(R^{**}) - R}{R} \qquad [28]$$

Only robust priors were evaluated for efficiency. Those with the tightest bound, such that efficiency was minimum, were considered most useful in the sense that the practitioner could use them confidently because they provide coverage with minimum overestimation of the population error rate.

III. Findings and Analysis

The research objectives, as previously stated in Chapter I, were to evaluate the reliability and validity of the FBP and to determine if any specific prior distributions improved the model effectiveness. A three-phased design was operationalized through use of a CDC Cyber 750 computer and programs, written in the Fortran language incorporating International Mathematical and Statistical Language (IMSL) subroutines. The methodology required the determination of the compatibility, robustness, and efficiency of selected priors. All thirteen prior distributions were compatible with the model. This chapter examines the robustness and efficiency of those priors relative to the selected study populations and draws conclusions concerning model reliability and validity.

Performance Trends

The Appendix, entitled "Performance Data," contains tables summarizing test results of each prior distribution for eight study populations. Four different error rates were examined for a relatively small population (1000), and a moderate-sized population (5000). These tables depict values for r**, R**, and the hypergeometric probability associated with each study population at various reliability levels. Further, they show the expected bound, E(R**), of the priors relative

to the test values. Before elaborating on the analysis of robustness and efficiency, some apparent trends in the test data were noted.

Effects on Minimum Estimated Population Errors (R**). The R** values represent the FBP minimum estimated number of errors in the population, as per Eq [23], for each possible number of observed errors, r. The performance tables in the appendix illustrate the increasing values for R** as the study population error rate and specified reliability increased. The relative R** values remained similar as population size increased. For example, as the population size increased from 1000 to 5000, the associated R** values also increased by approximately 500 percent. These trends were expected and indicate the adjustments the model makes for varying population sizes, reliability, and sample outcomes.

Effects on Hypergeometric Probability. The hypergeometric probabilities shown in the appendix were computed, as per Eq [24], and represent the likelihood of sample outcomes where the model provides coverage. These probabilities were compared with the specified reliability, as in Eq [25], to determine the robustness of the priors for the study populations.

In general, the hypergeometric probabilities decreased as the population error rate increased. That is, as the actual error rate increased, the FBP bound increased and, accordingly, a higher number of observed errors, r**, were needed to establish that bound. As such, the likelihood of these larger sample error outcomes grew more remote.

Within each study population the hypergeometric probability tended to increase as the reliability level increased. This trend was

directly attributable to the decreasing r** values associated with the increased reliabilities.

Generally, the hypergeometric probability for the larger population size was smaller than for the population size of 1000 for the same actual error rate and reliability. This trend did not affect the performance of priors across the population sizes.

The hypergeometric probability corresponding to a particular bound had to exceed the specified reliability for the priors to provide adequate coverage relative to the study populations. Those prior distributions supplying coverage were considered robust.

Robustness

Information from the Performance Data was consolidated into Tables III and IV, which depict robust prior distributions for the two population sizes. The tables reflect the expected bound, E(R**), computed from Eq [27], for the robust prior distributions. Additionally, those priors providing coverage for the particular population size illustrated in the table, but not for the other study population size, are noted.

Overall, the selected priors were robust for the study populations in slightly more than half the cases. With a population size of 1000, 108 of 208 possible cases (51.9 percent) were robust. For a population size of 5000, 105 of the 208 cases (50.5 percent) were robust.

Effect of Population Error Rate (u) Changes on Robustness. With few exceptions, when the actual population error rate exceeded the expected error rate, the model did not provide coverage and the prior

TABLE III

Robustness: E(R**) Values for N = 1000

		EXPECTED ERROR RATE												
		VERY LOW E(P) = .005			LOW E(P) = .01			MODERATE E(P) = .05			HIGH E(P) = .10			UNIFORM
VARIANCE:		Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	
AEP†	REL													
u=.005	.85	7.50	8.91		12.50	12.91		51.58	19.99		94.00	19.90		23.91
(5)	.90	7.51	10.32	11.30•	13.50	14.99	12.71	53.58	23.41	14.12	97.00	24.31		28.40
	.95	8.50	13.33	15.14	14.58	18.82	17.55	56.58	29.31	19.44	102.00	30.22	18.27	35.23
	.99	10.50	18.90	24.53	17.50	27.32	28.35	62.58	43.23	31.67	111.00	44.81	29.90	51.22
u=.01	.85				13.01			52.28	25.90		95.00	26.56		30.64
(10)	.90				14.00	18.92		54.28	29.64		98.00	31.22		35.56
	.95		16.38		15.26	23.29		57.28	36.22		103.00	37.87		42.95
	.99	11.00•	22.57	32.10	18.00	32.29	37.40	63.28	50.93	42.68	112.00	53.21	41.67	59.87
u=.05	.85							58.25			103.00	73.80		80.27
(50)	.90							60.25			106.00			86.76
	.95							63.25	84.49		111.00	90.42		97.41
	.99							69.59	103.69		120.00	111.10		119.50
u=.10	.85										113.00			136.83
(100)	.90										116.00			*
	.95									147.09•	121.00	148.37•	149.55•	158.15
	.99										130.29	*		183.98

• Coverage provided for N = 1000 but not for N = 5000.
* Coverage provided for N = 5000 but not for N = 1000.

† AEP = Actual Error Rate

TABLE IV

Robustness: E(R**) Values for N = 5000

		EXPECTED ERROR RATE												
		VERY LOW E(P) = .005			LOW E(P) = .01			MODERATE E(P) = .05			HIGH E(P) = .10			UNIFORM
VARIANCE:		Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	
AEP†	REL													
u=.005 (25)	.85	30.59	43.68		57.91	65.26		262.99	103.84		494.40	106.15		126.93
	.90	32.50	51.59	*	60.59	76.65	66.98	270.60	122.13	73.43	507.09	125.60		148.63
	.95	34.51	64.59	75.47	63.91	95.54	89.93	281.60	153.01	99.68	526.40	158.28	92.25	184.99
	.99	38.90	93.63	122.62	70.59	138.82	145.99	302.91	222.58	164.82	564.01	233.88	155.15	266.66
u=.01 (50)	.85				58.72			264.92	133.51		497.64	138.86		160.77
	.90				61.28	95.61		272.34	153.76		510.26	160.56		184.74
	.95		78.51		64.72	116.21		283.35	187.28		529.65	196.37		224.10
	.99	*	109.52	160.44	71.28	162.38	191.14	304.74	261.69	219.92	567.08	277.26	214.68	310.85
u=.05 (250)	.85							279.40			522.23	375.47		408.61
	.90							286.81			535.20			444.01
	.95							298.00	428.19		554.88	460.38		499.39
	.99							319.79	527.97		592.71	568.11		613.85
u=.10 (500)	.85										552.96			693.53
	.90										566.20			736.96•
	.95									*	586.28	*	*	804.15
	.99										624.85	877.87		938.65

• Coverage provided for N = 1000 but not for N = 5000.
* Coverage provided for N = 5000 but not for N = 1000.

† AEP = Actual Error Rate

distribution was not robust. When the prior expressed an anticipated error rate that equaled or exceeded the actual error rate, the priors were robust. For example, given a .01 population error rate and an anticipated error rate of .01, the prior distribution is robust for population error rates less than or equal to .01. However, given an anticipated error rate of .01 and an actual error rate of .05 or .1, the prior distribution is not robust.

This trend suggests some conclusions regarding model reliability and validity. The model generally produced robust priors when the population contained errors equal to or less than the amount anticipated. Yet, when the population error rate exceeded expectations, the model almost never generates robust priors.

The exceptions to this tendency partially accounted for robustness in slightly over 50 percent of the cases. At very low expected error rates (E(P) = .005) and higher reliability levels, the model did provide coverage for population error rates double those anticipated. This trend occurred at the REL = .99 level for all three different variances. This provides favorable support for model reliability and validity at these levels. For example, assuming an auditor specified a reliability of 99 percent and expected an error rate of one-half percent, this finding indicates that the FBP would provide coverage despite an actual error rate of 1 percent, or double that anticipated.

Effect of Population Size (N) on Robustness. Other trends are apparent when comparing Tables III and IV. With the exceptions noted by asterisks, the same coverage was provided for smaller and moderate-sized populations. Two cases were not robust at N = 1000 but were

robust at N = 5000; three cases were not robust at N = 5000 but were robust at N = 1000. Despite these minor deviations, the performance of priors seems to occur independent of population size.

Effect of Reliability Level (REL) on Robustness. In general, as the reliability level increased, increased robustness occurred. For example, with N = 1000, 21 of 108 (19.4 percent) were robust at the 85 percent reliability level. This compares to 34 of 108 (32.4 percent) at the 99 percent reliability level. The results are similar for the 5000 population size. Though not obvious at the small variance level, this trend became more apparent as the variance increased to moderate and high levels. As expected, an increasing expected bound, E(R**), accompanied the increases in reliability.

Effect of Prior Probability Variance (σ^2) on Robustness. Model coverage decreased as expected error rate variance increased from small to high. This deteriorating robustness occurred initially at the lower reliability levels and then moved to the moderate and high levels as the variance on the priors increased. Data in the Appendix illustrates the cause for this trend, which was mainly due to the lower hypergeometric probabilities associated with the high r** values needed at the lower reliability levels. Since these values did not exceed the specified reliability, the priors did not provide coverage.

As the variance in the priors increased along with the higher actual error rates, a trend was noted in values of the expected bound, E(R**). Initially, the E(R**) values tended to increase as the variance increased, for a given reliability level. This general trend existed for both very low and low expected error rates. Yet, as the

expected error rates increased to moderate and high, an apparent reversal occurred when tested against the lower actual error rates. Increasing E(R**) values for increasing variance continued when the expected error rate equaled or exceeded the actual error rate; however, when the actual error rate was less than that expected, the E(R**) values generally decreased (for the same reliability level) as the variance on the priors increased. This tendency becomes significant in the discussion of efficiency.

Uniform Distribution and Robustness. As discussed in Chapter II, the uniform distribution represents the non-informative prior implicitly assumed in classical sampling techniques. This assumed prior was tested to enable some conclusions regarding a comparison of classical techniques with the FBP model. As shown in Tables III and IV, this prior distribution performed adequately, with only one case where coverage was not provided (i.e., N = 5000, u = .10, REL = .90). This exception may have been partially due to roundoff error. The performance of this distribution, however, appears to detract from the model's utility. It would appear that using a lower expected error rate would be more effective; however, the research data indicate the uniform distribution is more robust. Despite this finding, the FBP utilizes smaller sample sizes than the classical approach.

Efficiency

As explained in Chapter II, efficiency measures the tightness of the expected bound relative to the actual number of population errors. The efficiency measure, given by Eq [28], indicates the average amount

the particular prior distribution will overestimate the number of actual population errors. Since auditors wish to minimize this overestimation while assuring coverage, efficiency becomes an important model characteristic.

Tables V and VI depict the efficiency values for robust priors. As with the tables of robustness, these efficiency tables are separated by population size. The values shown are expressed as percentages. Hence, a value of 50 implies that the expected bound overestimates the actual population error rate by 50 percent (i.e., efficiency = $\left(\frac{E(R^{**}) - R}{R}\right) \times 100$). Several trends evolve from an examination of these tables.

<u>Effect of Population Error Rate (u) on Efficiency</u>. As the actual population error rate increased for specific priors, efficiency improved. When the expected error rate equaled the actual error rate, efficiencies were relatively good, ranging from 10.6 percent to 391 percent. However, the efficiency of the prior distributions dramatically deteriorated as the difference between the expected error rate and the actual population error rate increased. For example, for a prior with an expected error rate of 10 percent and a small variance, the amount of overestimation was 2156 percent, or 539 errors, for a population of 5000 having only 25 actual errors.

<u>Effect of Population Size (N) on Efficiency</u>. The change in population size from 1000 to 5000 had relatively minor effects on the efficiency of priors. In general, the efficiencies were slightly better for the larger population at small variance levels, and slightly better for the smaller population at moderate and high variance levels,

TABLE V

Efficiency (%) for N = 1000

		EXPECTED ERROR RATE												
		VERY LOW E(P) = .005			LOW E(P) = .01			MODERATE E(P) = .05			HIGH E(P) = .10			UNIFORM
VARIANCE:		Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	
AEP[†]	REL													
u=.005 (5)	.85	50	78		150	158		932	300		1780	298		378
	.90	50	106	126	170	200	154	972	368	182	1840	386		468
	.95	70	167	203	192	276	251	1032	486	289	1940	504	265	605
	.99	110	278	391	250	446	467	1152	765	533	2120	796	498	924
u=.01 (10)	.85				30			423	159		850	166		206
	.90				40	89		443	196		880	212		256
	.95		64		53	133		473	262		930	279		329
	.99	10	126	221	80	223	274	533	490	327	1020	432	317	499
u=.05 (50)	.85							17			106	48		61
	.90							21			112			74
	.95							27	69		122	81		95
	.99							39	107		140	120		139
u=.10 (100)	.85										13			37
	.90										16			*
	.95										21	48	50	58
	.99										30			84

* Coverage provided for N = 5000 but not for N = 1000.

[†] AEP = Actual Error Rate

TABLE VI

Efficiency (%) for N = 5000

		EXPECTED ERROR RATE												
		VERY LOW E(P) = .005			LOW E(P) = .01			MODERATE E(P) = .05			HIGH E(P) = .10			UNIFORM
VARIANCE:		Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	Small	Mod	Large	
AEP†	REL													
u=.005	.85	22	75		132	161		952	315		1878	325		408
(25)	.90	30	106		142	207	168	982	389	194	1928	402		495
	.95	38	158	202	156	282	260	1026	512	299	2006	533	349	640
	.99	56	275	339	182	455	484	1112	790	559	2156	836	521	967
u=.01	.85				17			430	167		895	178		222
(50)	.90				23	91		445	208		921	221		270
	.95		57		29	132		467	275		959	293		340
	.99		119	221	43	225	282	509	423	340	1034	455	329	522
u=.05	.85							12			109	50		63
(250)	.90							15			114			78
	.95							19	71		122	84		100
	.99							28	111		137	127		146
u=.10	.85										11			39
(500)	.90										13			47
	.95										17			61
	.99										25	76		88

†AEP = Actual Error Rate

except at the very low expected error rates. For example, the larger population/small variance values overestimated the actual error rate by an average of 646 percent. In contrast, the small population/small variance values overestimated the actual error rate by a 743 percent average. On the other hand, the small population overestimated by an average of 276 percent and 301 percent for moderate and large variances, respectively, compared to an average of 289 percent and 344 percent for the same variances in the larger population. These differences, however, were not considered a weakness of the model.

Effect of Reliability (REL) on Efficiency. As expected, the efficiency values increased with increasing reliability levels, implying that the model consistently and increasingly overestimated the actual number of population errors as the desired confidence increased. This result followed from the previous observation where the expected bound increased as reliability increased. This trend highlights an apparent tradeoff between reliability and efficiency (efficiency tends to suffer at the expense of reliability).

Effect of Prior Probability Variance (σ^2) on Efficiency. As with the effects on robustness discussed earlier, changes in prior distribution variance seemed to result in a reversing trend on efficiency. Generally, the smaller variance level performed more efficiently than either the moderate or high variance level when the expected error rate equaled the actual error rate. When the expected error rate increasingly exceeded the actual error rate, the efficiency degenerated for all three variance levels. However, particularly for the moderate and high expected error rates, when the actual population error rate

was substantially lower than anticipated, the moderate and high variance levels performed more efficiently. For example, given a population size of 1000, REL = .99, and an expected error rate of 10 percent, if the actual number of population errors was only 5, versus the 100 anticipated, then the small variance level would result in 111 estimated errors, whereas the large variance level estimated 30 errors. Apparently, when the actual error rate is significantly less than the expected error rate, the moderate and high variances were markedly more efficient at the reliability levels for which they provided coverage.

Need for Tradeoffs

These findings seem to support the necessity for tradeoffs between robustness, efficiency, and reliability. For example, based on the values shown in Table V, if the auditor expects an error rate of 10 percent, he may receive better coverage by specifying a small variance, particularly when the actual population error rate approaches 10 percent. If, however, the actual population error rate is significantly less, specifying a moderate or high variance would seem to provide a much more efficient estimate of population error rate. Furthermore, at these higher variance levels, the model does appear to provide coverage, and fairly good efficiency, when the actual error rate approaches 10 percent, but only at the 95 percent reliability level.

It would appear that an "ideal" prior probability distribution was not found. That is, no prior was identified which provided coverage with good efficiency across all study populations. Instead, the

findings indicate the importance of specifying an expected error rate which closely matches the actual error rate. Further, analysis shows the apparent requirement to balance robustness, efficiency, and reliability. These considerations would become another of the subjective judgments the auditor must make when planning the audit.

The uniform distribution seems to emerge as the best overall candidate with regard to balancing the required tradeoffs. With near perfect robustness, this prior also performed efficiently, particularly with higher error rate populations. Though not especially encouraging, this finding would imply that, in the absence of a fairly strong and well-founded expectation of the probable error rate, a more conservative approach may be to specify a non-informative, uniform prior distribution.

Summary of Primary Findings

Several findings emerged from this study which appear relevant to the research objectives. These significant findings must, however, be tempered with an understanding of the methodology used in the research. They result from a practical analysis of the data, as opposed to a more rigorous statistical evaluation.

The primary findings were:

1. The beta prior distribution assumed in the model does not impose serious constraints on the selection of prior distributions, but does require some selection.

2. The model reliability and validity appears adequate, as evidenced by the robustness and efficiency of priors, when the actual error rate matches the expected error rate.

3. The model reliability and validity is suspect when the expected error rate significantly exceeds the actual error rate, because of the poor efficiency, and when the expected error rate underestimates the actual error rate, due to the lack of coverage.

4. The model provides coverage when the actual error rate equals or doubles an anticipated very low error rate at a high reliability level.

5. Population size does not significantly impact model performance.

6. Selection of priors with a small variance appears to provide better coverage across populations and more efficient performance when the actual error rate matches that expected, but moderate and high variances perform more efficiently when the expected error rate significantly overestimates the population error rate. These larger variances often do not, however, result in adequate coverage, particularly at lower reliability levels and higher actual error rates.

7. The uniform distribution is robust and relatively efficient across all study populations and reflects perhaps the most conservative prior distribution evaluated.

8. Tradeoffs between robustness, efficiency, and reliability seem a necessity when using the FBP.

IV. Conclusions and Recommendations

The findings presented in Chapter III provide some information to practitioners considering the use of the FBP and reveal several areas for continued study. This chapter examines the potential usefulness of the model and provides recommendations for future research.

Potential Use of the FBP

Given that the auditor can closely approximate the population error rate, the FBP could be incorporated into an efficient audit process, enabling reduced sample sizes, as demonstrated by Godfrey and Andrews (21:308-313). Using Tables V and VI, presented in Chapter III as a judgmental guide, the auditor could evaluate the possible tradeoffs and ramifications involved in the selection of particular prior distribution. Borrowing from Felix and Grimlund's suggestions for a similar model, the method of application would appear fairly simple: Assuming the auditor has access to an adequate computer facility, and algorithm incorporating the FBP could be developed. Once established, the operational use appears straightforward for a qualified auditor. The anticipated error rate and variance could be prespecified, along with the population size, materiality, and desired reliability. The computer could return a corresponding sample size and, after entering the number of errors detected, could return the probability of a material error. The material error's sensitivity to

prior judgment could then be tested through repeated applications of this procedure (19:39).

Perceived Shortcomings

The Finite Bayesian Procedure appears to have deficiencies. Several researchers address the shortfalls of attribute testing in the audit environment, emphasizing the need to estimate the dollar value of errors rather than the error rate. T. M. F. Smith states, "There is no obvious way of turning the error rate . . . into an upper limit for the money value of the errors" (32:269). Anderson and Teitlebaum state that "the methods which focus on error frequency seem to provide no meaningful conclusions in dollars" (32:269-270). These criticisms appear well-founded for substantive testing but inappropriate for compliance testing. As long as the compliance testing objective is to evaluate internal control systems guiding subsequent substantive testing, the need for quick and efficient methods such as the FBP should continue (20).

Other perceived shortcomings relate to the complexity of Bayesian models and their relative usefulness within the judgmental context of auditing. In part, this criticism apparently stems from general resistance to change, perhaps due to a lack of understanding. Bayesian methods, particularly the FBP, recognize the subjectivity of the audit environment and attempt to incorporate these judgments in a more objective fashion. The resultant benefits in smaller sample sizes also seem to warrant their continued consideration. The efficiency provided by the FBP appears even more appealing in light of the findings in

this study, where a non-Bayesian uniform distribution performed well. Coupled with the smaller sample sizes, the FBP seems to provide a more advantageous approach than the classical technique. Computational complexities could be overcome with increased use of computer procedures.

Additional possible weaknesses with the FBP concern prior probability distributions elicitation. The assessment of priors remain a major stumbling block to the use of Bayesian techniques (21:313). Though this study provides some insight into the performance of specific priors against selected populations, it also highlights the need for an accurate and verifiable means of eliciting auditor's subjective beliefs. The model's inability to estimate population error rates greater than those specified seems to be a major weakness. This apparent failure does not tolerate errors in auditor judgment during the assessment process and deserves more scrutiny. The accounting profession seems to recognize this shortcoming and research literature continues to address it (7-10; 13; 14; 15; 18).

Another value requiring predetermination by the auditor using Bayesian methods is materiality. Again, judgment influences the auditor's choice. Though current research aims at determining the effects of auditor decisions in this regard, more work is required (15:589).

Recommendations for Future Research

This study represents an analysis of selected prior probability distributions and populations in the evaluation phase of the FBP;

however, there seems to be opportunities for future research.

An appropriate area for continued effort seems to be in statistically evaluating the results of this study. The analysis phase of this research centered on practical implications. Specifically, the analysis focused on the robustness and efficiency of selected priors, drawing implications concerning the model's reliability and validity. Future research could possibly use analysis of variance (ANOVA) and regression to test relationships of the variables and provide more objective conclusions regarding prior probability distributions and model reliability and validity. Future efforts could focus on the tradeoffs among robustness, efficiency, and reliability, presenting the information in more easily comprehensible form, such as in tables or on graphs.

Research design variation could provide additional information on model performance. For example, changes in sample size may impact the reliability and validity of the FBP. Also, testing the model against actual populations, rather than the hypergeometric distribution, may increase external validity of research results.

Moreover, an extensive study examining the range of priors between an expected error rate of one-half percent and 5 percent appears to be justified. As noted previously, a serious drawback in the model exists because of its general inability to provide coverage at error rates greater than those expected. However, the coverage provided at double the expected rate for the very low error rates, as shown in Tables III and IV, reverses this trend and may indicate that the FBP does tolerate auditor misjudgment under certain circumstances. Consequently, more

research in this area would seem relevant.

Another area for study is the reversal in efficiency as the variance on the priors increases. Moderate and large variance priors seem to provide more efficient coverage when the actual and expected error rate differ significantly. Small variance priors perform more efficiently when the actual and expected error rates match. An in-depth evaluation of this tendency may supply useful information with which auditors can temper their judgment in prior distribution selection.

The possible benefits in smaller sample sizes accruing from the planning phase of the FBP were not examined in this study. Continued research in this area could result in more persuasive evidence on the usefulness of the model, particularly through a comparative study with classical techniques. The uniform distribution appears as a reasonable starting point from which to extend this work, and emphasizing the different efficiencies may prove enlightening. Further, by focusing on the elicitation technique proposed by Godfrey and Andrews, future study could expand the knowledge base of subjective assessment research.

Appendix: Performance Data

Very Low Expected Error Rate - Small Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	7	1.0000	7.50	25	0	30	1.0000	30.59
	.90		0	7	1.0000	7.51		0	32	1.0000	32.50
	.95		0	8	1.0000	8.50		0	34	1.0000	34.51
	.99		0	10	1.0000	10.50		0	38	1.0000	38.90
.01	.85	10	3	10	.0692		50	16	50	0	
	.90		3	11	.0692			15	51	0	
	.95		2	10	.2637			13	51	0	
	.99		0	10	1.0000	11.00		9	50	0	
.05	.85	50	*	*	0		250	*	*	0	
	.90		*	*	0			*	*	0	
	.95		*	*	0			*	*	0	
	.99		*	*	0			*	*	0	
.10	.85	100	*	*	0		500	*	*	0	
	.90		*	*	0			*	*	0	
	.95		*	*	0			*	*	0	
	.99		*	*	0			*	*	0	

* A large number of sample errors was needed to generate a bound, resulting in a zero probability.

Priors: $r' = 99.495$ $E(P) = .005$
$n' = 19899$ $\sigma^2(P) = .00000025$

Very Low Expected Error Rate - Moderate Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	6	1.0000	8.91	25	0	31	1.0000	43.68
	.90		0	7	1.0000	10.32		0	38	1.0000	51.59
	.95		0	10	1.0000	13.33		0	50	1.0000	64.59
	.99		0	15	1.0000	18.90		0	77	1.0000	93.63
.01	.85	10	1	12	.6531		50	1	57	.6377	
	.90		1	14	.6531			1	66	.6377	
	.95		0	10	1.0000	16.38		0	50	1.0000	78.51
	.99		0	15	1.0000	22.57		0	77	1.0000	109.52
.05	.85	50	9	50	.0531		250	11	267	.0107	
	.90		9	54	.0531			10	264	.0268	
	.95		8	54	.1166			9	268	.0611	
	.99		6	53	.3834			7	271	.2324	
.10	.85	100	21	104	.0004		500	23	500	.0001	
	.90		20	104	.0011			22	503	.0003	
	.95		18	101	.0068			21	516	.0007	
	.99		16	104	.0320			18	517	.0093	

Priors: $r' = .99$ $E(P) = .005$
$n' = 198$ $\sigma^2(P) = .000025$

Very Low Expected Error Rate - Large Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	1	15	.4102		25	1	76	.3973	
	.90		0	5	1.0000	11.30		1	91	.3973	
	.95		0	8	1.0000	15.14		0	40	1.0000	75.47
	.99		0	16	1.0000	24.53		0	80	1.0000	122.62
.01	.85	10	1	15	.6531		50	1	76	.6377	
	.90		1	18	.6531			1	91	.6377	
	.95		1	23	.6531			1	115	.6377	
	.99		0	16	1.0000	32.10		0	80	1.0000	160.44
.05	.85	50	5	52	.5731		250	5	254	.5657	
	.90		5	56	.5731			5	278	.5657	
	.95		4	54	.7567			4	270	.7450	
	.99		3	58	.8944			3	292	.8842	
.10	.85	100	11	101	.0077		500	12	531	.2964	
	.90		11	107	.0077			11	524	.4167	
	.95		10	108	.0214			10	532	.5499	
	.99		8	108	.1166			8	537	.7966	

Priors: $r' = .24375$ $E(P) = .005$

$n' = 48.75$ $\sigma^2(P) = .0001$

Low Expected Error Rate - Small Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	12	1.0000	12.50	25	0	57	1.0000	57.91
	.90		0	13	1.0000	13.50		0	60	1.0000	60.59
	.95		0	14	1.0000	14.58		0	63	1.0000	63.91
	.99		0	17	1.0000	17.50		0	70	1.0000	70.59
.01	.85	10	0	12	1.0000	13.01	50	0	57	1.0000	58.72
	.90		0	13	1.0000	14.00		0	60	1.0000	61.28
	.95		0	14	1.0000	15.26		0	63	1.0000	64.72
	.99		0	17	1.0000	18.00		0	70	1.0000	71.28
.05	.85	50	5	*	0		250	*	54	0	
	.90		*	*	0			*	*	0	
	.95		*	*	0			*	*	0	
	.99		*	*	0			*	*	0	
.10	.85	100	*	*	0		500	*	*	0	
	.90		*	*	0			*	*	0	
	.95		*	*	0			*	*	0	
	.99		*	*	0			*	*	0	

* A large number of sample errors was needed to generate a bound, resulting in a zero probability.

$r' = 98.99$ $E(P) = .01$

$n' = 9899$ $\sigma^2(P) = .000001$

Low Expected Error Rate - Moderate Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	9	1.0000	12.91	25	0	47	1.0000	65.26
	.90		0	11	1.0000	14.99		0	57	1.0000	76.65
	.95		0	14	1.0000	18.82		0	74	1.0000	95.54
	.99		0	22	1.0000	27.32		0	114	1.0000	138.82
.01	.85	10	1	17	.6531		50	1	84	.6377	
	.90		0	11	1.0000	18.92		0	57	1.0000	95.61
	.95		0	14	1.0000	23.29		0	74	1.0000	116.21
	.99		0	22	1.0000	32.29		0	114	1.0000	162.39
.05	.85	50	6	51	.3834		250	7	276	.2324	
	.90		6	55	.3834			6	277	.3839	
	.95		5	54	.5731			5	264	.5657	
	.99		3	51	.8944			3	251	.8842	
.10	.85	100	14	101	.1121		500	15	510	.0705	
	.90		13	100	.1873			14	509	.1216	
	.95		12	102	.7102			13	520	.1961	
	.99		10	104	.5550			10	502	.5499	

Priors: $r' = .98$ $E(P) = .01$

$n' = 98$ $\sigma^2(P) = .0001$

Low Expected Error Rate - Large Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	1	18	.4102		25	1	91	.3973	
	.90		0	5	1.0000	12.71		0	29	1.0000	66.98
	.95		0	9	1.0000	17.55		0	47	1.0000	89.93
	.99		0	18	1.0000	28.35		0	95	1.0000	145.99
.01	.85	10	1	18	.6531		50	1	91	.6377	
	.90		1	21	.6531			1	109	.6377	
	.95		1	27	.6531			1	138	.6377	
	.99		0	18	1.0000	37.40		0	95	1.0000	191.14
.05	.85	50	4	51	.7567		250	4	254	.7450	
	.90		4	56	.7567			4	280	.7450	
	.95		3	53	.8944			3	265	.8842	
	.99		2	55	.9692			2	280	.9642	
.10	.85	100	9	100	.6910		500	10	542	.5499	
	.90		9	106	.6910			9	530	.6814	
	.95		8	107	.8079			8	534	.7966	
	.99		6	105	.9512			6	529	.9442	

Priors: $r' = .2375$ $E(P) = .01$

$n' = 23.75$ $\sigma^2(P) = .0004$

Moderate Expected Error Rate - Small Variance

		N = 1000 n = 100					N = 5000 n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	51	1.0000	51.58	25	0	261	1.0000	262.99
	.90		0	53	1.0000	53.58		0	269	1.0000	270.60
	.95		0	56	1.0000	56.58		0	280	1.0000	281.60
	.99		0	62	1.0000	62.58		0	301	1.0000	302.91
.01	.85	10	0	51	1.0000	52.28	50	0	261	1.0000	264.92
	.90		0	53	1.0000	54.28		0	269	1.0000	272.35
	.95		0	56	1.0000	57.28		0	280	1.0000	283.36
	.99		0	62	1.0000	63.28		0	301	1.0000	304.74
.05	.85	50	0	51	1.0000	58.25	250	0	261	1.0000	279.40
	.90		0	53	1.0000	60.25		0	269	1.0000	286.81
	.95		0	56	1.0000	63.25		0	280	1.0000	298.00
	.99		0	62	1.0000	69.59		0	301	1.0000	319.79
.10	.85	100	*	*	0		500	*	*	0	
	.90		*	*	0			*	*	0	
	.95		*	*	0			*	*	0	
	.99		25	100	0			*	*	0	

* A large number of sample errors was needed to generate a bound, resulting in a zero probability.

Priors: $r' = 94.95$ $E(P) = .05$

$n' = 1899$ $\sigma^2(P) = .000025$

Moderate Expected Error Rate - Moderate Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	14	1.0000	19.99	25	0	73	1.0000	103.84
	.90		0	17	1.0000	23.41		0	89	1.0000	122.13
	.95		0	22	1.0000	29.31		0	117	1.0000	153.01
	.99		0	35	1.0000	43.23		0	181	1.0000	222.59
.01	.85	10	0	14	1.0000	25.90	50	0	73	1.0000	133.51
	.90		0	17	1.0000	29.64		0	89	1.0000	153.76
	.95		0	22	1.0000	36.22		0	117	1.0000	187.28
	.99		0	35	1.0000	50.93		0	181	1.0000	261.69
.05	.85	50	4	59	.7567		250	4	300	.7450	
	.90		3	54	.8944			3	274	.8842	
	.95		2	50	.9692	84.49		2	256	.9642	428.19
	.99		1	52	.9955	103.69		1	267	.9944	527.97
.10	.85	100	8	100	.8079		500	8	502	.7966	
	.90		8	107	.8079			8	538	.7966	
	.95		7	107	.8953			7	540	.8853	
	.99		5	105	.9811			5	533	.9773	

Priors: r' = .9 $\quad$ E(P) = .05

n' = 18 $\quad$ σ^2(P) = .0025

Moderate Expected Error Rate - Large Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	1	20	.4102		25	1	105	.3973	
	.90		0	5	1.0000	14.12		0	27	1.0000	73.43
	.95		0	9	1.0000	19.45		0	47	1.0000	99.68
	.99		0	19	1.0000	31.67		0	102	1.0000	164.83
.01	.85	10	1	20	.6531		50	1	105	.6377	
	.90		1	24	.6531			1	125	.6377	
	.95		1	31	.6531			1	159	.6377	
	.99		0	19	1.0000	42.68		0	102	1.0000	219.92
.05	.85	50	4	59	.7567		250	4	299	.7450	
	.90		3	52	.8944			3	266	.8842	
	.95		3	61	.8944			3	312	.8842	
	.99		2	64	.9692			2	328	.9641	
.10	.85	100	8	105	.8079		500	8	529	.7966	
	.90		7	101	.8953			7	511	.8853	
	.95		6	100	.9512	147.09		6	508	.9442	
	.99		5	108	.9811			5	554	.9773	

Priors: $r' = .1875$ $\quad E(P) = .05$

$n' = 3.75$ $\quad \sigma^2(P) = .01$

High Expected Error Rate - Small Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	93	1.0000	94.00	25	0	491	1.0000	494.40
	.90		0	96	1.0000	97.00		0	504	1.0000	507.09
	.95		0	101	1.0000	102.00		0	523	1.0000	526.40
	.99		0	110	1.0000	111.00		0	561	1.0000	564.01
.01	.85	10	0	93	1.0000	95.00	50	0	491	1.0000	497.64
	.90		0	96	1.0000	98.00		0	504	1.0000	510.26
	.95		0	101	1.0000	103.00		0	523	1.0000	529.66
	.99		0	110	1.0000	112.00		0	561	1.0000	567.08
.05	.85	50	0	93	1.0000	103.00	250	0	491	1.0000	522.23
	.90		0	96	1.0000	106.00		0	504	1.0000	535.20
	.95		0	101	1.0000	111.00		0	523	1.0000	554.88
	.99		0	110	1.0000	120.00		0	561	1.0000	592.71
.10	.85	100	4	101	.9942	113.00	500	2	504	.9997	552.96
	.90		2	100	.9998	116.00		0	504	1.0000	566.20
	.95		0	101	1.0000	121.00		0	523	1.0000	586.28
	.99		0	110	1.0000	130.29		0	561	1.0000	624.85

Priors: $r' = 89.9$ $E(P) = .10$

$n' = 899$ $\sigma^2(P) = .0001$

High Expected Error Rate - Moderate Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	13	1.0000	19.90	25	0	72	1.0000	106.15
	.90		0	17	1.0000	24.31		0	89	1.0000	125.60
	.95		0	22	1.0000	30.22		0	118	1.0000	158.28
	.99		0	36	1.0000	44.81		0	187	1.0000	233.88
.01	.85	10	0	13	1.0000	26.56	50	0	72	1.0000	138.86
	.90		0	17	1.0000	31.22		0	89	1.0000	160.57
	.95		0	22	1.0000	37.87		0	118	1.0000	196.37
	.99		0	36	1.0000	53.21		0	187	1.0000	277.26
.05	.85	50	3	52	.8944	73.80	250	3	264	.8842	375.47
	.90		3	57	.8944			3	293	.8842	
	.95		2	53	.9692	90.42		2	273	.9642	460.38
	.99		1	54	.9955	111.10		1	285	.9944	568.11
.10	.85	100	8	107	.8079		500	8	541	.7966	
	.90		7	104	.8953			7	525	.8853	
	.95		6	103	.9512	148.37		6	524	.9442	
	.99		5	112	.9811			4	507	.9926	877.87

Priors: $r' = .8$ $\quad E(P) = .10$

$n' = 8$ $\quad \sigma^2(P) = .01$

High Expected Error Rate - Large Variance

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	1	20	.4102		25	1	103	.3973	
	.90		1	24	.4102			1	123	.3973	
	.95		0	7	1.0000	18.27		0	35	1.0000	92.25
	.99		0	16	1.0000	29.90		0	86	1.0000	155.15
.01	.85	10	1	20	.6531		50	1	103	.6377	
	.90		1	24	.6531			1	123	.6377	
	.95		1	31	.6531			1	158	.6377	
	.99		0	16	1.0000	41.67		0	86	1.0000	214.68
.05	.85	50	4	60	.7567		250	4	302	.7450	
	.90		3	53	.8944			3	268	.8842	
	.95		3	62	.8944			3	315	.8842	
	.99		2	64	.9692			2	331	.9641	
.10	.85	100	8	107	.8079		500	8	539	.7966	
	.90		7	102	.8953			7	519	.8853	
	.95		6	101	.9512	149.55		6	516	.9442	
	.99		5	110	.9811			5	563	.9773	

Priors: $r' = .125$ $E(P) = .10$

$n' = 1.25$ $\sigma^2(P) = .04$

Uniform Distribution

		N = 1000, n = 100					N = 5000, n = 100				
u	REL	R	r**	R**	PROB	E(R**)	R	r**	R**	PROB	E(R**)
.005	.85	5	0	17	1.0000	23.91	25	0	92	1.0000	126.93
	.90		0	21	1.0000	28.40		0	111	1.0000	148.63
	.95		0	27	1.0000	35.23		0	144	1.0000	184.99
	.99		0	42	1.0000	51.22		0	220	1.0000	266.66
.01	.85	10	0	17	1.0000	30.64	50	0	92	1.0000	160.77
	.90		0	21	1.0000	35.56		0	111	1.0000	184.74
	.95		0	27	1.0000	42.95		0	144	1.0000	224.10
	.99		0	42	1.0000	59.87		0	220	1.0000	310.85
.05	.85	50	3	57	.8944	80.27	250	3	291	.8842	408.61
	.90		2	50	.9692	86.76		2	257	.9642	444.01
	.95		2	58	.9692	97.41		2	303	.9642	499.39
	.99		1	61	.9955	119.50		1	316	.9944	613.85
.10	.85	100	7	104	.8953	136.83	500	7	527	.8853	693.53
	.90		7	111	.8953	145.18		6	507	.9442	736.96
	.95		6	111	.9512	158.15		5	503	.9773	804.15
	.99		4	107	.9942	183.98		4	549	.9926	938.65

Priors: $r' = 1$ $\quad E(P) = .5$

$n' = 2$ $\quad \sigma^2(P) = .288675$

Bibliography

1. AICPA, Auditing Standards Executive Committee. "Codification of Auditing Standards and Procedures," Statement on Accounting Standards No. 1, AICPA, New York, 1973.

2. Arkin, Herbert. Handbook of Sampling for Auditing and Accounting. New York: McGraw-Hill Book Company, 1984.

3. -----. Sampling Methods for the Auditor. New York: McGraw-Hill Book Company, 1982.

4. Bailey, Andrew D., Jr. Statistical Auditing: Review, Concepts, and Problems. New York: Harcourt Brace-Jovanovich, Inc., 1981.

5. Birnberg, J. "Bayesian Statistics: A Review," Journal of Accounting Research, 2(16): 108-116 (Spring 1964).

6. Blocker, E. "A Multivariate Variables Extension of Bayesian Statistical Sampling Methods for Auditors," AIDS Proceedings, 13-15 (October 1977).

7. Chesley, G. R. "Subjective Probability Elicitation Techniques: A Performance Comparison," Journal of Accounting Research, 16(2): 225-241 (Autumn 1978).

8. -----. "Subjective Probability Elicitation: The Effect of Congruity of Datum and Response Mode on Performance," Journal of Accounting Research, 15(1): 1-11 (Spring 1977).

9. -----. "The Elicitation of Subjective Probabilities: A Laboratory Study in an Accounting Context," Journal of Accounting Research, 14(1): 27-48 (Spring 1976).

10. -----. "Elicitation of Subjective Probabilities: A Review," The Accounting Review, 44(2): 325-337 (April 1975).

11. Committee on Basic Auditing Concepts. "A Statement of Basic Auditing Concepts," The Accounting Review, Supplement to Vol. 47: 15-74 (1972).

12. Comptroller General of the United States. Standards for Audit of Governmental Organizations, Programs, Activities, and Functions. Washington DC: Government Printing Office, 1972.

13. Corless, J. C. "Assessing Prior Distributions for Applying Bayesian Statistics in Auditing," The Accounting Review, 47(3): 556-566 (July 1972).

14. Crosby, Michael A. "Bayesian Statistics in Auditing: A Comparison of Probability Elicitation Techniques," The Accounting Review, 56(2): 355-365 (April 1981).

15. -----. Implications of Prior Probability Elicitation on Auditor Sample Size Decisions," Journal of Accounting Research, 18(2): 585-593 (Autumn 1980).

16. Cyert, R. M. and H. Justin Davidson. Statistical Sampling for Accounting Information. Englewood Cliffs NJ: Prentice-Hall, Inc., 1962.

17. Ericson, W. A. "Subjective Bayesian Models in Sampling Finite Populations," Journal of the Royal Statistical Society, 2: 195-224 (1969).

18. Felix, W. L., Jr. "Evidence on Alternative Means of Assessing Prior Probability Distributions for Audit Decision-Making." The Accounting Review, 51(4): 800-807 (October 1976).

19. ----- and R. A. Grimlund. "A Sampling Model for Tests of Composite Accounts," Journal of Accounting Research, 15(1): 23-41 (Spring 1977).

20. Godfrey, James T., Associate Professor, Accounting. Personal interviews. University of Georgia, 18 April through 20 April 1984.

21. ----- and Richard W. Andrews. "A Finite Bayesian Model for Compliance Testing," Journal of Accounting Research, 20(2): 304-315 (Autumn 1982).

22. ----- and John Neter. "Bayesian Bounds for Monetary-Unit Sampling in Accounting and Auditing," Audit Research Working Paper Series, Report No. 83-001. University of Georgia, 1983.

23. Huber, Peter J. Robust Statistics. New York: John Wiley and Sons, Inc., 1981.

24. Ijiri, Y. and R. A. Leitch. "Stein's Paradox and Audit Sampling," Journal of Accounting Research, 18(1): 91-107 (Spring 1980).

25. Johnson, Kenneth P. "Evaluating Internal Control," Accountant's Handbook (Sixth Edition), edited by Lee J. Seidler and D. R. Carmichael. New York: John Wiley and Sons, Inc., 1981.

26. Kraft, W. H., Jr. "Statistical Sampling for Auditors: A New Look," Journal of Accountancy, 126(8): 49-56 (August 1968).

27. McRae, T. W. Statistical Sampling for Audit and Control. New York: John Wiley and Sons, 1974.

28. Meek, Gary E. and Stephen J. Turner. Statistical Analysis for Business Decisions. Boston MA: Houghton-Mifflin Company, 1983.

29. Meigs, Walter B. and others. Principles of Auditing. Homewood IL: Richard D. Irwin, Inc., 1973.

30. Neter, John and James K. Loebbecke. "Behavior of Major Statistical Estimators in Sampling Accounting Populations: An Empirical Study," AICPA Auditing Research Monograph 2, 1975.

31. Scott, W. R. "A Bayesian Approach to Asset Valuation and Audit Size," Journal of Accounting Research, 11(2): 304-330 (Autumn 1973).

32. Smith, T. M. F. "Statistical Sampling in Auditing: A Statistician's Viewpoint," The Statistician, 28(4): 267-280 (1979).

33. Sorensen, J. E. "Bayesian Analysis in Auditing," The Accounting Review, 44(1): 555-561 (January 1969).

34. Tracy, J. A. "Bayesian Statistical Methods in Auditing," The Accounting Review, 44(1): 90-98 (January 1969).

35. Winkler, Robert L. An Introduction to Bayesian Inference and Decision. New York: Holt, Rinehart and Winston, Inc., 1972.

VITA

Captain Stephen P. Marino was born 21 February 1953 in Port Chester, New York. He graduated from high school in Glastonbury, Connecticut, in 1971. He received a Bachelor of Science degree in International Affairs from the United States Air Force Academy in 1975. He was a Distinguished Graduate from Undergraduate Navigator Training at Mather AFB CA, and served as a KC-135 instructor navigator from 1976 to 1980. He was an instructor navigator in the RC-135S at Eielson AFB AK until entering the School of Systems and Logistics, Air Force Institute of Technology, Wright-Patterson AFB OH, in May 1983.

Permanent Address: 1611 George Washington Drive
Dayton, Ohio 45432

VITA

Captain Kevin O'Shaughnessy was born 5 February 1955 in Haddonfield, New Jersey. He graduated from high school in 1973, and from the United States Air Force Academy in 1977, earning a Bachelor of Science degree in Management and the distinction of Honor Graduate. After receiving his commission in the USAF, he entered pilot training at Williams AFB AZ and received his wings in July 1978. He then served as a C-141 pilot and flight instructor in the 18th Military Airlift Squadron, McGuire AFB NJ, until entering the School of Systems and Logistics, Air Force Institute of Technology, Wright-Patterson AFB OH, in May 1983.

Permanent Address: 442 Tavistock Boulevard
Haddonfield, New Jersey 08033

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GSM/LSY/84S-21	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Systems and Logistics 6b. OFFICE SYMBOL (If applicable) AFIT/LS	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) Air Force Institute of Technology Wright-Patterson AFB, Ohio 45433	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO
11. TITLE (Include Security Classification) See Box 19	
12. PERSONAL AUTHOR(S) Stephen P. Marino, B.S., Capt, USAF	Kevin O'Shaughnessy, B.S., Capt, USAF
13a. TYPE OF REPORT MS Thesis 13b. TIME COVERED FROM ____ TO ____	14. DATE OF REPORT (Yr., Mo., Day) 1984 September 15. PAGE COUNT 84
16. SUPPLEMENTARY NOTATION	Approved for public release: IAW AFR 190-17. Lynn E. Wolaver 14 Sept 84 LYNN E. WOLAVER Dean for Research and Professional Development Air Force Institute of Technology (ATC) Wright-Patterson AFB OH 45433

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	
12	01		Auditing, Statistics, Bayes, Bayes Theorem

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Title: AN ANALYSIS OF ROBUST AND EFFICIENT PRIORS ASSOCIATED WITH A FINITE BAYESIAN MODEL FOR COMPLIANCE TESTING

Thesis Advisor: Jeffrey J. Phillips, Major, USAF

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Jeffrey J. Phillips, Major, USAF	22b. TELEPHONE NUMBER (Include Area Code) 513-255-4845 22c. OFFICE SYMBOL AFIT/LSY

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

This research examines the reliability and validity of the Finite Bayesian Procedure model through an evaluation of robust and efficient prior probability distributions. The model, developed by James Godfrey and Richard Andrews, presents a different approach to compliance testing in auditing. This study utilizes small and moderate-sized populations, four population error rates, a fixed sample size, and four reliability levels. In addition, four expected error rates, based on a beta prior probability distribution and ranging from very low to high, combined with three variance levels and a uniform distribution, are used to evaluate the model. The results indicate that the model is adequately reliable and valid. However, the uniform distribution seems to perform best of all prior probability distributions tested. Moreover, tradeoffs between robustness, efficiency, and reliability seem a necessity when using the Finite Bayesian Procedure model.

END

FILMED

1-85

DTIC